Summary

About this Release

There is very strong demand from analysts, particularly within government and universities, to access micro-data collected by agencies, such as the Australian Bureau of Statistics (ABS), for the purpose of developing and evaluating policy. To help meet this demand, the ABS is planning to develop a remote server which would automatically return the output from remotely submitted statistical programming code. In allowing such access, the ABS is legally obliged to ensure that any information (e.g. analysis output) it releases is not likely to enable the identification of the particular person or organisation to which it relates. This paper considers the problem of managing the disclosure risk associated with releasing analysis output, including regression parameters and model diagnostics for generalised linear models, by a remote server. While this paper restricts attention to surveys where all variables are categorical, these variables can be defined without restriction. The disclosure risk is managed by adding noise in two different ways. The first adds noise to the input data prior to analysis and the second adds noise to the counts present in the estimation equation. All inferences using the statistical output released by the server are valid in the presence of adding noise. The methods are evaluated using the 2008 National Health Survey. The results show that perturbing counts in the estimating equation leads to a very small loss in accuracy.